古籍数字化与传统诗学的范式革新

李 飞 跃

(清华大学人 文学院, 北京 100084)

摘　要: 随着诗歌电子文本的各种聚合,产生了大量宏文本和超文本;分词、标注及语义技术的发展,也使诗歌知识向细粒度、高维度发展。数据库作为一种文本,让诗歌经典秩序重置,诗学概念也面临从实体到关系的再定义。基于大数据技术的整体性与过程性分析,改变了经验研究的随机性与无标准,让诗学重新作为一种技艺,增强了科学性与合法性;同时,模型和算法也让隐变量、暗知识和潜语境浮现,凸显了文本的互文性与非定域,让人在意义链条中的作用更加凸显。古籍数字化、知识增强和多模态语言大模型技术的深入发展,引发了诗学从经典重塑、概念重构、关系重建到价值重估的系统变革。

关键词: 数字人文;诗学技艺;互文性;范式

中图分类号:G 250.7　　**文献标识码:**A　　**文章编号:**1000-5919(2024)06-0088-12

古籍电子化是对抄刻本的复现,但又基于数字技术让原来的线性结构呈现为立体化的超文本,为知识的获取、标注、表示、阐释、取样与比较等带来了根本性变革。数据库、知识库和语言大模型作为一种文本,一方面形成了超越传统尺度的宏文本,另一方面又因切词、标引而形成多种语义单元和知识碎片。检索界限消失后,古籍数据库可以最大限度地一站式获取所需资料。“将科学的严格、系统、明确、程序的方法特征带到人文学科中,来解决那些迄今为止大多以偶然的方式被处理的人文学科问题”①,走出辑佚式考证和举例式论证,逐渐形成了过程可重复、数据可验证、方法可复用、结论可推广的科学研究范式。利用文本挖掘、社会网络、时空分析等方法,可从非结构化文本中提取信息和知识,揭示文本背后隐藏的变量和模式。标记、连接与模型分析,又生成了系统化空间文本和交互界面,让各种知识碎片重新聚合,原有场景、活动和过程逐渐消融。文献形态和技术方法的变革,不可避免地引发了对传统诗学研究中的一些基本问题的重审和反思。

一、基于数据库文本的经典秩序重塑

数据库作为众多文本的集合,不仅是文本数量的叠加,更是结构功能的不同。基于数字文献形态的检索、聚合与智能问答,是基于算法和概率的选择而非主观撷取或随机抽样。我们每次检索阅读时,都在与这些算法进行交互②。基于大数据技术和全样本全过程的研究中,以往常用的抽样方法开始显出狭窄、随机的一面。数字化让离散知识、非量化知识和未被现有学科框架包含的知识显现出来,让原有史实变得不再确定:“如果我们的文学史写作是围绕着‘重要的’作家进行的,那么我们就必须问一问他们是什么时候成为‘重要作家’的,是什么人把他们视为‘重要作家’,根据的又是什么样的标准。”③

收稿日期:2024-04-12

作者简介:李飞跃,清华大学人文学院长聘教授。

① 苏珊·霍基,葛剑钢译:《人文计算的历史》,《文化研究》2013 年第 4 期,第 173—193 页。

② Long, Hoyt and Richard Jean So, “Literary Pattern Recognition: Modernism between Close Reading and Machine Learning”, *Critical Inquiry*, Vo1. 42. No. 2, 2016, p. 236.

③ 宇文所安著,田晓菲译:《他山的石头记——宇文所安自选集》,南京:江苏人民出版社 2003 年版,第 8 页。

分类和排序是经典建构的主要方式,利用度、率、量、值及参数(指标)、模型、算法等,可将研究对象和问题量化,将定性问题转化为定量问题,变得有标准可依。当王兆鹏等通过量化方式对唐诗、宋词排行,也引发了强烈质疑:“唐诗是情感的艺术表达,阅读者有各自的爱好,要找出大家都能接受的名篇排行,本身就是几乎不可能的任务。王兆鹏团队,‘尝试用统计学的方法来衡量测度公众的关注度和名篇指数’,当然是花气力而不易讨好的努力。但其方法,则是西方的、现代的。按照古人的办法,其实找几位名家,喝茶沐浴之余,提出个名单,也可以很权威。”①唐诗、宋词排行榜所做的是将历史上千百位“名家”的评价赋值,提出名单取个位与千位数的商值虽同,代表性与置信度却不同。随着样本量的剧增,根据中心极限定理,经典文本的平均值会呈正态分布,能更客观地反映诗词名篇的影响。

唐诗宋词排行榜其实重新界定了什么是经典、什么是影响力,王兆鹏认为作品影响力可从三个层面来衡量:“一是对创作者的影响,包括引用、化用、仿效、改编、翻译等,体现出作品的典范性和启示性;二是评点者的评论和学者的研究,反映出作品在文学批评、学术研究层面的美誉度和关注度;三是在普通读者中的传阅度和知晓率。”在这些参数基础上,充分考虑结构权重,计算出各篇作品的综合得分。定性研究讲究特殊深入和主观经验,更多地依赖于研究者的体验与感悟。它与定量研究之间并非不兼容,“同样具有审美性、艺术性、观赏性的体育比赛项目如艺术体操、花样滑冰、高台跳水等,都可以通过主观评量来打分,比优劣、定高下,为什么文学作品不可以用量化标准、数据来评价其艺术质量呢?”,相反“因为它们有一套完备的评价指标体系和操作规则,裁判按章办事,所以评分结果相对公平公正”。②

测量逐步排除不确定性,真实反映了对象属性和特征差异,让感觉评判变得有标准。罗忼烈曾指出,宋词界“大家”与“名家”的区别,是以“古今著名的词选到底选哪些人的作品最多”为标准,并以《花庵词选》《草堂诗余》等选录南北宋二十四家词数为例,给出了数据说明。③ 同样,王兆鹏和刘尊明对宋代词人历史地位的定量分析,从存词、版本、品评、研究文献、历代词选、当代词选的篇数对宋代 30 位代表词人做出统计后排列名次,考察了“词人地位与词作数量的关系”“词评家与词选家的异同”“古今变异与本世纪词学研究的走向”等问题。④ 每一个阶段和结果都可以明确表示出来,甚至发现事物之间有共同的数学关系,建立适当的数学模型并逐步加以修改,以对事物状况、趋势和相互关系所做的陈述更具普适性。

约克斯指出:“21 世纪的文学学者不能再满足于轶闻式的证据,不能再从那些少量的、即使可以称为‘代表性’的文本中得到随机的‘事实’。我们发现的那些有意思的东西,我们必须努力将它们放置在文本的上下文中去理解,这些上下文包括大量‘不那么有意思’的文本。”⑤数字化文本让数量和维度激增,研究变量的各种状态或变异,让批评研究立足于整体性而非抽样推断。以往研究者易将某一或某些例证所反映的现象普遍化,研究数据“只是它经典化了的一小部分作品,它们甚至占不到已出版文学的百分之一”,“除非文学评论家也看了另外 99.5%的作品,否则就无法理解那 0.5%的经典化作品”,但是又没办法让学者们去细读那些“大量未读作品”。⑥《唐诗排行榜》让我们重新思考什么是诗人的影响力、如何才算经典,需要综合考虑哪些要素,包括选本、媒介、评点等指标和权重,而不仅是某一群体的经验和好恶。

建模是理解大样本的一种方式,它关注“整体”与“结构”,让知识在更大尺度上被测量,测量才能被

---

① 陈尚君:《唐诗凭什么排名》,《东方早报》2012 年 2 月 19 日。

② 王兆鹏:《古典文学研究数据化的可行性》,《文学遗产》2022 年第 6 期,第 4—9 页。

③ 罗忼烈:《试论宋词选集的标准和尺度》,《文艺理论研究》1983 年第 4 期,第 110—114 页。

④ 王兆鹏,刘尊明:《历史的选择——宋代词人历史地位的定量分析》,《文学遗产》1995 年第 4 期,第 47—54 页。

⑤ Jockers, Matthew L. Macroanalysis: *Digital Methods and Literary History*, Urbana-Champaign: University of Illinois Press, 2013, p. 8.

⑥ Moretti, Franco *Distant Reading*, London: Verso, 2013, pp. 65-71.

真正理解和认识。每个变量的属性和特征都能以一个数值来表示,形成指标体系,从而变得相互通约和可优化迭代。用文本相似度、新颖度代替以往基于主观感受的风格评判,能让更多文本具有可比性。特别是可视化技术让原来的时间性文本具有了空间化特征,这一全局、全知、全貌的研究视角,本质上是一个“更多特征维度、更细知识粒度的目录或索引”①,进而能够揭示文本内部或文本之间的隐含内容和潜在语义联系。把海量文本数据转换为静态图像或动态图形,通过交互手段控制数据展示,使隐藏的数据关系和现象可直观呈现,也比以前更容易看到节点更新和趋势性内容。指标化与可视化,让我们的思考过程变得可见与可操作。

面对被研究对象的多样性和多变性,引入模型可以通过参数、变量的调整和算法的优化不断接近事实真相。研究同一个现象的不同侧面,会得到完全不同的数学模型,特征量的取值大小和变化足以表达所研究对象的主要状态及演变趋势。如基于唐诗生成的两种近体诗律模型,为律绝创作和诗体研究提供了基本准则。通过对《全唐诗》全部五七言四八句诗歌的格律统计发现,启功、林庚先生比王力先生所提格律模型应用率相对较高,但整体上完全符合标准模型的诗作比例较小,只是随着拗变度的增加而覆盖面益广。通过考察两种诗律模型的历史演变可见,模型的确立不仅建立在自发的创作实践基础上,还受诗法理论和诗学观念影响而渐趋自觉。② 排行榜作为一种模型,定类、定序、定距、定比等测量包含了主观与客观成分,实质是对传统文献评价方法的细化,同时改进了传统研究中全面性、客观性不足及证据缺口的问题。

数据本身是客观的,但设置的指标及权重却是主观的。模型中每个因素作用方式不同,导致计量结果并不唯一。同一组数据,从不同角度进行分析可以传达两个完全相反的事实,每个人都可以根据自己的立场并通过对于数据的分析得到想要的结果。悖论是对一个理论系统而言的内部逻辑矛盾,任何一个悖论都相对地被包含在某个理论体系中,属于一定的历史范畴,反映了人的认识的历史的局限性和相对性。悖论并非就是错误,反而孕育了研究范式的变革和新的解决路径,如“落体悖论”导致了牛顿力学的诞生,“以太悖论”为相对论的创立开辟了道路。量化可以让信息缺失而产生的悖论在一定范围内被解决,并让人们认识到哪些因素是如何导致我们形成了如许看法,从而寻求更为周密合理的答案。利用模型和算法也让人文研究可以被测量、验证和重复操作,一些争议无须“翻烧饼”或推倒重来,而是致力于某些要素或关系的度量优化,获得可验证与可累积的普遍知识。

二、诗学概念的量化与结构化重置

学术史可以说是一部概念串联起来的文化史,古典诗学中的大量概念是具有语境特征和不精确性,需要通过不断辨析和校准来把握对象和形成共识。变量是具有可变化特征的因素,或是对某个一特征的测量。面对“概念所揭示的事实到底存不存在”“其特征和定义是什么”“是如何演变的”等问题,学界往往会有很多争议,需要从上下文语境中来领会。如“风格”这一概念,“既可以涵盖时代、地域、流派、个人等诸多方面,又可以涉及社会、文化、历史和心理等诸多层面,同时又是一个逻辑上有丰富内涵、解释力比较广泛,又具有相当弹性的概念”③。落实到具体研究中,就需要进一步明确其内涵、共同特征及测量标准,事实上是对概念的重新定义。数字化将以往被分裂和隔绝的事物在细粒度上重新对齐和连接,改变了我们对文献、文本、知识的认识路径和把握尺度。

诗歌声律统计中,样本的选择、前置标准与统计方法都可能影响结论的显著性与准确性。如“四声

---

① 王军:《从人文计算到可视化——数字人文的发展脉络梳理》,《文艺理论与批评》2020 年第 2 期,第 18—23 页。

② 李飞跃:《唐诗格律的统计分析及问题》,《文学遗产》2022 年第 5 期,第 44—56 页。

③ 张婷:《作为艺术史核心概念的“风格”的概念史——艺术史知识建构的方法论探索》,《文艺理论研究》2020 年第 5 期,第 208—218 页。

八病”“齐梁格”“律体”“拗体”等概念的界定与验证,不仅基于诗句中四声的分布及组合关系,还关系到四声字的数量与比例。一些诗律特征可能是基数与概率使然,而非诗人的自觉探索或刻意创新。早期韵书中,平声字约占 40%,上声字 18%、去声字 18%、入声字 24%,那么,“x 平 xx 入”句式的概率应为 9.58%,而出现“x 上 xx 去”的概率仅为 3.24%。如果要计算遵守永明体“蜂腰”二五异声的概率,可用所有句子分别去除二五字同平、上、去、入的比例,其自然合格率为 72%;反之,28%为犯“蜂腰”的自然病犯率。这样就可得到一个合律度:H=1-[P(实际病犯率)/(自然病犯率)]%。这一公式考虑了不同规则自然病犯率的差异,体现了诗人实际病犯情况与随机取字下病犯情况的比值。① 当然,科学定律不总能用一个特定的数学公式来表达,但却可用模型和算法来把握事实整体及其变化过程。

统计方法一般只能验证实体性的,不能否定生成性。我们不能先验地假设存在一个确定的诗律公式,以避免循环论证。无论抽样还是举例论证,按照标准模型区分古近体,又依据其中的律诗来验证是否合律,筛选出来的诗歌当然全都合律。唐代诗律究竟为何,是一种最大公约数,还是各要素联动的矩阵函数?统计结果打破了唐诗中的后世所谓之格律标准,那么我们今天所遵循的标准格律又在何时确立?诗律不是单一模型,而是包含多种变体的模态。流传下来的唐代诗集,也存在被规范或再加工的可能。一些别集整理与语言学家的研究,通过是否合律来推导和判定唐诗的读音、异文、句法、节奏及韵律等,又不免堕入新的循环,让唐诗格律变成一种自洽自足的诗学建构。② 文学史上的一些著名命题,如“文学自觉说”“唐宋变革论”等都是标准多样但可量化的命题。如围绕“文学自觉”说的争议,从先秦到隋唐的不同时期判定,以及从作者、作品、文论、文体等方面建模,可从整体上揭示这一命题的丰富内涵和应用场景。

诗歌和日常生活中有大量模糊性语言,好坏、美丑、善恶、虚实、真假等用语并不具备一成不变的内涵,而是以个人经验、直觉和所处情境赋值。随着知识颗粒度的细化和向量化,原来二元化的命题已无法很好反映事实整体,真与假、美与丑、好与坏、善与恶,甚至存在与虚构、物质与精神都展现了巨大的中间地带。二元化的定性为更多层级的量化定义所改写,如“亮度”对开关、“颜值”对美丑的重新定义。用定量的方式来把握事物的本质,可以变节点为量度而实现一种精确把握。凡是可以测量的,甚至凡语言可以表达的,就是显性的或确定的,意味着能被人感知和度量,进而就可以复制、转译并加以描述。虽然对事情的描述不一定是事情本身,但多个描述的叠加与交互验证,能更好地接近和揭示真相。度、量、值、率等计量单位在文本分析中的应用消解了二元论,增强了研究的客观性、精确性与可操作性,既可以降低抽样数据的稀疏性和偏差度,又可增加动态过程的可见度,缓解了实体与属性、现象与本质、是然与或然之间的紧张对立,让等值换算和比较成为可能。

传统概念隐喻的是实体指代和关系描述,而数字化文献的概念则需要深入要素及结构。20 世纪 40 年代,金岳霖就提出任何一个概念都有结构,结构就是概念里各要素之间单箭头或者双箭头的联系。为双箭头或者单箭头赋值的时候,这个值可以是变量,可以是参量。只要有“结构”,就可以有“模型”,如它们的权重是多少,相互之间是什么关系。当概念有了结构的时候,统计学家和数学家就可以帮上忙了。③ 我们以往习惯的实体与本质是具有情境和历史依赖的,社会或自然不可能一劳永逸的决定本质,只能在追随行动者的过程中历史性地考察本质。我们应避免从有固定界限的概念出发去把握事物全貌和事实变化,悬置一些条件和变量而仅用部分条件特征来描述不同时代、群体和视域中的文学体式、风

---

① 基于这一理念和统计结果,“蜂腰”可作重新界定,并刻画其随时代发展而呈现的不同趋势与规则。见陆泉宇:《永明体到近体句内声调对立规则的嬗变——以“蜂腰”与“二四异声”为代表》,《数字人文》2023 年第 1 期,第 154—179 页。

② 李飞跃:《唐诗格律的统计分析及问题》,《文学遗产》2022 年第 5 期,第 44—56 页。

③ 汪丁丁:《社会科学定量分析方法基础探讨》,《社会科学战线》2010 年第 1 期,第 40—46 页。

格及意象。易言之,新的概念定义涵盖应然性、或然性与生成性,能够展现其历时性及共时性上的互动结构。

主流、共识往往是后来居上的理解,动态的数字表征系统让我们可以更好认识和把握同一文本的不同向度及其时代嬗变。数字化概念作为一个历史的动态系统被识别和表征,基于数值、量度来进行更具针对性和建设性的探讨,可以避免画地为牢、自说自话。以诗体研究为例,至少包括共时、历时与比较三个层面。一种诗歌文体或流派到底存不存在,需要建立文本库、对照组和特征模型来验证而非仅凭抽样阅读或先验范式。基于语料库通过多种算法提取和生成关键词群,以发现文类的指标和关键词的关联度,确证其能作为一个类型的独特性。以往是通过要素、特征和范畴等进行描述,现在还需考虑包括哪些要素、各要素权重是什么、它们是一个什么样的结构关系。

概念定义的嬗变是学科发展的常态,一个概念的活力往往是它的含混与多变,因为“一个概念或关键词的重要性或关键发展,常常不在于频繁使用,而是取决于被论辩、被争夺的强度,或在某个历史时期和关键时刻的多义性和争议性及影响力,或在观察和解释社会、政治状况时的不可或缺”①。在概念重新界定之后,再对其在历史上的变化区间进行定位分析,构建知识模型,通过调节参数而把握和呈现其演进路线、阶段特征。从不同侧面建立对照组,确保概念的区分度。有实体、有关系、有过程也有对照,才能避免概念的含混和临文定义。

概念的量化界定也难免存在一些问题。其一,定量研究的推论或结论都来自不完整的测量和数据,数据搜集很难完美无缺。其二,“均值”“变量”过于简化和系统性偏差,不具有超越普遍性的效能。其三,从理论假设到变量、指标存在技术偏误与落差,有限变量、线性相关耦合而构成的理论模型无法呈现更多不确定的非线性关系。其四,当涉及证明某个“因”会带来一定的“果”,就必须面对内生性问题,即因变量或自变量变化的原因,如遗漏变量、自选偏误、样本偏误、联立偏误等。其五,词云、主题模型等方法将文本中的高维信息压缩到较低维度的数字上,便于统计和对比,也更具概括性和解释性。但是,把对文本内容的处理简化为向量运算,以空间上的相似度表达语义的相似度,这种方法毕竟是低维度概念的呈现。超越概念名词和关键词的功能性定义,尤其需要从实体转向关系,对更多隐变量进行把握和刻画。

三、从实体到关系与隐变量的涌现

究竟是先有关系还是先有关系者,“似乎总是先有关系者才有关系,但细细探究,便会发现有些内在关系在逻辑上先于关系者,其中最基本的就是差异关系:没有差异,也就无法断定和辨认关系者”②。关系也是一组变量,随着研究的深入,从不同要素、角度和阶段挖掘,会让更多变量涌现。自变量和因变量之间的线性关系毕竟少数,更多是非线性关系。隐变量、暗知识是容易被忽略的知识类型,就像我们体内的微量元素,数量和比重极少却是不可或缺。由于存在大量无法感受或表达描述的暗知识,每个人都掌握了其他人所不具备的分散知识,它们可能是不被已有框架容纳甚至相互矛盾的,却是整体的重要构成部分。

以往我们对事物的本质特征的认识往往是从实体和要素出发,但信息时代关系的权重日益增大,各种网链概念的流行反映了关系对实体的重新定义。概念界定已从原来的实体本位向关系本位转变,更关注要素结构及系统功能。“当我们看一个系统时,往往只会注意到系统的要素,而忽略掉系统的连接和功能。实则“只要不触动系统的内在连接和总目标,即使替换掉所有的要素,系统也会保持不变,或

---

① 方维规:《什么是概念史》第八章,北京:三联书店 2020 年版,第 320 页。

② 郦全民:《计算与实在——当代计算主义思潮剖析》,《哲学研究》2006 年第 3 期,第 82—89 页。

者只是发生缓慢的变化。相反,如果内在连接改变了,系统就会发生巨大的变化。”①越来越多的分散知识和未被现有学科框架所包含的实践知识,都需要我们更多从交叉、跨界等关系的维度去考察和把握。“事实”都是在某些先验假设和不同范畴化视角下的产物,背后有其整体性的信念、价值和假设。

作为一种语义网络,知识图谱旨在组织并可视化知识,可以灵活地对文本中的实体、概念、属性及其关系建模。建立可比较的量化标准,如等类、等序、等距和等比测量标准,可以动态立体地对知识进行测量,考察诗歌在流传中的区间意义和跃迁。数字化文本具有极强的关联性与查询能力,随着词向量技术尤其上下文敏感的词向量技术的进步,文本将在多维空间中投射意义。一些机构开源了中文词向量数据和语料库,不仅有助于分析古典文献的结构与影响,也有助于激活文本中更多的意义节点。标引和连接让类型化特征增强而特异性减弱,聚合分析和聚类研究就在改变它们作为离散性知识的属性与功能。

随着数字化语料库中更多知识来源的显现,可见一些著名诗人善于熔铸经史、剪裁锦文,一些名著也体现了深层的累积性特征。化用诗句就是《红楼梦》诗词创作的一个重要方式,而文本相似度是判定化用的重要标准。我们利用数字人文方法的“编辑距离”“词袋模型”“doc2vec”与“LDA 主题模型”,筛选出《红楼梦》所化用的唐宋诗句及其诗集来源,分析它们在词句、篇章层面的关系,考察曹雪芹的阅读经验与《红楼梦》诗意风格的生成关系。文本相似度技术能更为全面客观地给出《红楼梦》诗歌化用诗句的情况,直观反映《红楼梦》诗歌和前人诗歌基于文本要素上的关联,不仅能与前人研究所得相互验证,也有超出一般阅读经验的新发现。通过诗词文本的溯源,不仅可以看出《红楼梦》的构思与立意,也可考见曹雪芹的知识结构及来源,有助于我们从更丰富的维度理解《红楼梦》的文本生成及文化背景。② 同样,李杜诗歌与《文选》之间的关系被更多发现和揭示。虽然他们都曾学《文选》和拟古诗,但李白与杜甫却形成了截然不同的风格,确立了两种典范。知识的溯源无损于诗人或作品的伟大,反而让我们看到伟大是何以产生,认识到什么是真正的艺术创造及其关键变量。

传统的诗歌史或文学史通过保留重要作家作品,剔除凡作,描述了一个连续性累积进步的图景。如果将整体放大到这些一般作品,这种连续性和进步性不免要被打断,经典成为多样性与可能性之一种。最有意义和价值的不是结果,而是导致这种结果的一连串变化。每个文本都可以看作是不同文本的关键词按照不同权重混合的产物,量化研究把研究对象中那些可以客观测量的方面给揭示出来,忽略了无法客观测量的方面。量化研究需要大样本数据, 但真正随机的大样本并不容易获得。数字文本的标记,实体、属性等识别和取值并不唯一确定,而具有天然随机性。统计中某个前提条件发生变化,就可能得出完全不同的结论。近体诗律作为一种韵文标准模型,如果考虑到诗歌文本的异文、多音字、声调与语调、方音、韵书(切韵、唐韵、广韵、平水韵)、诗集整理等参数和变量,其客观性与确定性将被降解。随着知识粒度的细化,变量和维度随之增多,统计带来的可能不是确证,而是以更多参数和算法为表征的思辨。

数学语言是一种明确、统一和高度形式化的语言,不会因使用者的身份、经验和角度而不同。但一些诗学命题本是歧义、多变,可检验性差,较难达成共识和得出确定性结论。在超文本、多模态环境中,不再是由歌者到听者或作家到读者的单向传播,而是一种读者变成联合作者、文本相互映射的回路。这让任何语言学上的表达都只是对意义的部分近似,难以达成清晰无歧义的结论。计量方法一方面让“人本身所具有的不同于其他物种的本质属性被抹煞,一个个鲜活的个体身上最重要、最独特的方面被

① 德内拉·梅多斯:《系统之美:决策者的系统思考》第一章,邱昭良译,杭州:浙江教育出版社 2012 年版,第 25 页。

② 李飞跃、宋佳霏:《基于文本相似度分析的〈红楼梦〉化用唐诗研究》,《中外文论与文化》2023 年 54 辑,第 213—228 页。

忽略了,而次要的、一般的方面却被保留在了模型当中”①;另一方面,研究者的某些主观意愿、偏好的量化极难,数据的取舍、分类、定义和赋值都需要由人来掌控。确证偏见、便利抽样、虚假联系等现象,让人们前所未有地处于盲人摸象般的困局,也给诗歌研究带来了学科化危机。

诗学文献的数字化有助于增强客观性,同时又展现出巨大的未知与可能性。一方面是发现和验证新的知识,另一方面也引发更多争议和悖论。数字化让知识的线性结构和平面范域及其规则被打破,让离散知识系统化、隐性知识显性化,形成了一种新的四通八达的高维知识网络。超文本和多媒体技术带给文本一个个探索之谜,让原来基于线性叙事的历史时间模糊,文本矩阵变成了一个空间场域。超文本、多媒介与超媒体等将多层代码作为阅读内容,提供的是一种沉浸式与交互式的全息诗学。边界拓展、粒度细分和关系测度也改变了原有的整体性,让以往所有知识都呈现出一种碎片化和即时性特征。非线性叙事、虚拟场景、动态交互等特征,也构成了对传统诗学理论的颠覆式挑战。

四、数字化技艺与诗学的合法性

合法性是指正当性、权威性和实际有效性,源于重建一般性知识与公共知识,数字化技术让原来的碎片化、去中心化的知识具有通约和重新整合的可能。从诗歌文本细读到文化研究的拓展,越来越难以形成共识。生成式理论和过程性方法不断去中心化,文化研究的琐碎和技术化带来了深刻地范式危机。资料考订、字词解释和背景关联,总能找到适合的证据,也总能有异见甚至左见。“言之成理即可”,人言言殊,释放各式观点和看法的同时,也消解了共识,削弱了人文研究的权威性与合法性。众多新兴学科建立之初经常借助统计方法来强调客观实在和提升学术地位,统计学和数学凭借其符号的抽象化、结构化表达而被视为科学的象征。甚至一些传统学科如经济学、政治学、社会学、心理学、传播学等也纷纷借助统计学,开始了计算转向,以在信息化时代强化其科学性与合法地位。

亚里士多德说“技艺摹仿自然”②,技艺是一种共同的创作法则和实践路径。诗歌作为一种技艺,最初是就创作而言,语言学转向让诗歌的细读批评也成为一种技艺。技艺派甚至主张“技艺是所有诗歌和诗人之间的最大公约数,是所有诗人共通共享的,可以通过对技艺的讨论,互相促进,共同受益。而生活则无从谈起,因为每个人有自己不同的生活,不同的生活态度,价值取向;谈论生活就会陷入意识形态的争论”③。技艺具有中立性、客观性与可操作性,不仅是诗人的工具、文本的规则,也是读者的门径,是诗歌专业化、特殊性的本质因素。技术实践规则的建立,有助于在同一标准上重建共识。

文献数字化与相关技术的进步推进了文本分析尤其形式研究的迅速发展。从文献外部特征转向内部知识单元、从简单频次分析转向语义特征分析、从文献计量转向多学科交叉研究,字符、词量、词类、词性、段落、句法等特征都可统计,甚至计算词义演变、关键词共现、平均段落长度等。引入特征变量可以把离散知识整理成系统、抽象成模型,进而实现数理统计和分析。诗词的属性数据如“一首诗词的作者、标题、诗体、词体、题材、意象、时间、空间,一个作家的姓名、字号、郡望、籍贯、出生地、生卒年、家庭成员、朋友同僚、社会身份,甚至一首格律诗的韵字、对仗,一阕词的词牌、词格”④,都可以为量化数据提供计量基础。新技术转变为新方法,通过控制一系列变量,使不同类别具有可比性。通过对实体、属性等标引,可以进行反事实研究和模拟分析,发现已有研究的忽略变量与或然趋势。

---

① 许加明、陈友华:《数据质量、前提假设与因果模型——社会科学定量研究之反思》,《社会科学研究》2020 年第 2 期,第 130—139 页。

② 也译作“技术模仿自然”,参见亚里士多德:《物理学》第二章,张竹明译,北京:商务印书馆 1982 年版,第 48 页。

③ 雷武铃:《与新诗合法性有关: 论新诗的技艺发明》,《江汉学术》2013 年第 5 期,第 40—46 页。

④ 邵大为:《你知道“唐诗排行榜”哪首排第一? 诗词大数据揭示的有趣现象》,“道中华”微信公众号 2022 年 12 月 30 日。

语言的可计算性研究,促进了近年的作品形式、文本生成和文体研究的繁荣。向量空间模型特征权重计算方法、潜在语义分析技术和聚类知识图谱,可计算单双音节词、词长、句长等语言特征出现的比例,以及共现词、独有词、实词和虚词使用频率和词汇丰富度,让作者判断、文体计量、语义网络、人物关系和文本主题分析成为新的技艺。正则表达式可以帮助我们从全文数据中提取具有特殊写法的信息,句式句法和格律统计分析等让诗歌语言及体式研究的纵深度与科学性得以提升,也延续了“俄罗斯形式主义、风格批评、结构主义,并给予它们应有的实证基础”①。数字技术改进了诗歌研究的可测量与标准化,让文本分析的预设、程序和结果具有一致性与可重复性。散点图、图表、网络、公式等让关系明晰和秩序强化,也提升了诗歌批评的技术门槛。数字化和编码化带来的结构化使各种关系之间可以比较,并通过不同场景下产生的新指标而相互佐证,使系统地揭示与组织知识成为可能。

诗歌文献的规则性与通约性,是引进科学方法的基础。包弼德指出:“古代汉语是一种非常有规律的语言,古代汉语文本甚至比英语文本更适合运用‘文本挖掘’技术提取信息。”②量化方法与模型分析,与基于语言规则的文本分析传统相合,进一步拓展了研究方法和手段。模型通过不同的表示方法,可以做出真正有代表性的归纳,对研究对象的内部结构和关系作出整体揭示。学术科范化、精英化的发展,也需要建立更加精确的评估方法与指标体系。数字技术的应用,已不仅是从文本中寻求客观性,而通过技术手段对其进行客观性改造。

通过标注和连接,让各个实体和节点显现,把对象的复杂性从简单的程序中以可见方式演化出来。孙茂松教授团队基于 Bert 语言模型在 90 余万首中国古典诗词数据上训练并开放了中国古典诗词的预训练模型 BERT-CCPoem,利用该模型能够得到每个诗句 512 维的向量表示③,并可采用夹角余弦公式进一步计算出向量之间的距离即语义相关度。这为我们考察诗词知识来源及相似度提供了技术支持,即不仅从字词层面,还能基于语义层面精准溯源。程宁通过分析古诗中的“沾衣之泪”和“凄怆之悲”话题网络,发现早期诗歌中的相关话题不断重复并演变为固定句式出现在诗的尾联。以《古诗十九首》“泪下沾裳衣”和阮籍《咏怀》“凄怆伤我心”为代表的句式,更是在唐代诗歌中被反复使用。④ 这与宇文所安对诗歌创作的判断十分接近,即“有必要从‘诗歌材料’入手,把任何一个特定的文本都视为共享的材料库存之一小部分的具体实现,而不是独立的‘创作’”⑤。这在一定程度上弱化了诗人和作品的独立性,而更关注诗歌的整体性关联与制作技艺。

文献数字化和大数据中的“整体”已发生改变,细节阐释的边界无所在又无所不在。引入更多维度和参数后,测量结果不仅取决于文本,更取决于语境,以及测量方式和各要素的测量组合。每一个字词的含义根据上下文关系而变化,它们都在一定条件和语境下概率性成立。每个字词的含义都受其他字词出现与否及排列方式的影响⑥,测量结果可能取决于我们如何进行测量,或者我们选择了什么样的测量组合。数字化技术让诗歌技艺得以可见与可操作,诗歌不再是难以分解、具有独立审美价值的艺术品,而成为可以被不同艺术技巧重新结构、适应不同传播格式和受众审美偏好的超文本。

诗歌研究作为一门学科知识,主要通过技艺来确立其客观性与合法性。数字化技术的引入,尤其统

---

① 弗朗哥·莫莱蒂,张靖译:《定量研究之前途和问题:我的个人反思》,《数字人文研究》2023 年第 1 期,第 18—23 页。

② 纪明、曾江:《信息技术为中国历史研究带来新机遇》,《中国社会科学报》2010 年 9 月 17 日第 3 版。

③ 词向量或词嵌入将词汇映射到高维向量空间中,使得相似或相关的词在空间中的距离较近。GitHub 开源地址:https://github.com/THUNLP-AIPoet/BERT-CCPoem。

④ 程宁:《数字人文视域下〈文选〉与唐诗的互文空间》,《清华大学学报(哲学社会科学版)》2023 年第 6 期,第 75—88 页。

⑤ 宇文所安:田晓菲校:《中国早期古典诗歌的生成》,胡秋蕾等译,北京:三联书店 2012 年版,第 16 页。

⑥ N-gram 模型的核心思想就是“近朱者赤,近墨者黑”。

计和可视化工具的使用,重建了诗学研究的技艺门槛。在人文学科领域,定量方法“无疑能限制那种按照一般概念和‘大致估计’去作武断推论的可能性,排斥那种在没有准确定义的情况下在经典著作的片言只语中驰骋无度的作风,防止滥用不准确的概念和词句去掩盖浅薄和发挥偏见以至于攻击科学的现象再度发生”①。以计算的方法来研究文学和文化,可以增强研究的客观性、权威性与合法性,也“不是要用数学严谨性取代学者们数百年来发展出的阐释技巧的虚拟人文学科。它是增强版的人文学科,在最好的情况下,能展现最仔细的细读读者往往也看不见的新类型的证据和仔细考虑过的理论观点,二者联手产生新的批评研究”②。正是“增强版的人文学科”之中所要增加的部分提升了作为职业的文学研究的“文化权威”和“文化合法性”③。

技术工具的引入,一定程度上弥补了人的感官在观察、检验事实时的不足。同时,学术研究尤其诗歌分析无法脱离个人经验,数据收集、算法设计尤其意义阐释都是基于人的主观能动。初始条件、运行程序和参数算法是取决于人而非机器本身,“所有事无巨细的人文阐释,都是和算法环环相扣的;所有的阐释和应用无不是关于算法的阐释应用,不理解内置的模型和算法,就无法在计算批评的过程中真正地运用工具,更谈不上有针对性地根据模型来设计实验、研究问题了”,否则我们无法相信,“一种仅仅基于关键词计算出来的‘情绪曲线’,便代表了小说家、诗人的情感走向;在不经任何统计检验的情况下,就可以言之凿凿地用于一系列批评生发”。④

五、数字诗学的互文性与非定域

时空是我们用以描述一个事物如何影响另一个事物,及其相互作用的局限性的设定结构。我们习惯基于定域实在的经典系统来分析语义,以为按不同顺序或时空测量同一体系的同一对象,结果总是相同的。非定域性在信息时代的数据库和知识库中,语义关联不因距离变大而削弱,甚至是没有或超越距离的自成系统。基于语言文本和书籍形态的二维字符串组合,具有较强的定域性,从语言文本的封闭性、词语的循环阐释到版权制度,都划定了较强的界域。甚至如麦克卢汉所说,印刷术和书籍文化成为主导的形式,“在文学与艺术、艺术与生活之间竖立起了一道又一道的围墙”⑤。通过词语阅读来理解诗歌,信息和体验受制于文字符号。数字化文本具有多模态特征,可以融括文字、图像、声音、色彩等多种表意符号。数字文献中的多层代码,形成了一种文本层压,其理解是作者、文本和读者的意见聚合,整体上构成一种理解和阐释的开放性和多元性场域。数字化诗歌将词语与声音、图像跨媒介结合,以书本无法实现的多模态输出,阅读阐释被交互体验所取代。

互文原指在上下文中相关的语义单位参互见义,如诗歌创作和解释中的“无一字无来历”。现代互文性指文本之间的渗透和转换关系,不仅有“互相指涉、互相映射”的同质性,还有异质性和对话性。数字文献中互文性不仅是指主观上对文本有意的引用、改写,也指客观上两个文本之间的可能性关联,如基于词汇、语义、惯例、事件或任何可能的关联。这种关联让文本兼具含义和意义两个层面,含义可能保持不变,但意义不断变化。互文让线性文本变而为立体网络,“互文性的特点在于,它引导我们了解一

① 宋健:《社会科学研究的定量方法》,《中国社会科学》1982 年第 6 期,第 97—105 页。

② 霍伊特·朗,苏真,安德鲁·派博等:《推进计算文学研究——对笪章难〈以计算的方法反对计算文学研究〉一文的讨论》,《山东社会科学》2019 年第 8 期,第 40—55 页。

③ 姜文涛:《作为一种文学研究方法的数字人文——印刷文化基础设施,20 世纪文学批评史,以及文学社会学》,《中国比较文学》2019 年第 4 期,第 29—47 页。

④ 赵薇:《作为计算批评的数字人文》,《中国文学批评》2022 年第 2 期,第 157—166 页。

⑤ 埃里克·麦克卢汉著,弗兰克·秦格龙编,何道宽译:《麦克卢汉精粹》,南京:南京大学出版社 2000 年版,第 452 页。

种新的阅读方式,使得我们不再线形地阅读文本”①。表现形式的多样性和阅读开放性,让语汇和文本的意义不仅出乎其内,更是出乎其间。

定域实在认为物体自身内在地拥有一些明确的、界定清晰的本质特征,测量结果由隐变量决定。隐变量的考察一般是在定域内进行,但对于“经典”或“影响”等概念而言,无法通过发掘更多隐变量来定义,虽然每个隐变量都会对排行产生影响。不断发掘隐变量,也存在效度问题。首先,隐变量是无法穷尽,也无法恰当刻画,尤其无法准确找出根本原因。相较本体论层面的可量化,方法论和认识论层面的可量化更复杂也更难达成共识。诗歌文本的校勘、注释、翻译、引用等活动,均会形成文本间关系,带动发生语境和语义变化。其次,不同人基于不同预设和范畴的测量数值不同,不具有可重复性和可验证性。因为测量是基于已有事实,而“已有”随着时空及方法等不断发生变化。最后,量化研究无法确定变量之间的因果关系。大数据有无限多角度和维度发生知识链接与交叉验证,更多表现为相关性而非因果性。有时起决定作用的可能是更大的未知,这是一定条件下无法理解和认知的,当然也就谈不上测量和量化。

非定域和互文性也是量子力学的根本特性,它主张粒子之间的强相关性不可能用定义每个粒子各自属性的定域隐变量来解释,纠缠对中包含的信息必须在粒子间非定域共享。非定域带来语义纠缠和叠加状态,参与者所处语境起决定作用。不同的语境,立场、角度与方法不同,对事物的评判完全不同。文学史教材中对侍从诗人和台阁体的评价较低,也可能是我们的解码方式问题。处于高层政治生活的文人,对具体信息和文字都有高度敏感性,不可能采取直抒胸臆和张扬个性的表达方式,而只能是把情感加密,通过典故、写景和陈词滥调来掩抑真情实感。说什么、不说什么、怎么说,都有心迹和微表情可推究,信息含量不同于日常化写作,不能因为解码方法和解读方式无法抵达而斥为无病呻吟。明清文字狱就是政治高压下的过敏反应,朱元璋长期的民间生活经验让他对诸如隐语尤其双关、借代、析字、藏词等十分熟稔也比较敏感,这种生活经验和思维方式再加上当时政治情势,也就能够解释他何以借诗文而行挞伐,大兴文字狱来铲除皇权威胁。如果我们仅从语言文学或者文学艺术的角度,对审美、情感和艺术性欠佳的作品弃如敝屣,结果就是失去了一类诗歌的解读能力。

在一般参照系中,存在因果关系的两事件都是因前果后,互文性和非定域则会导致特殊的因果结构。当存在多个测量角度和逻辑推理路径时,历史顺序与读者获取知识信息的顺序都会影响到对事物因果关系的感受和判断。解释一个字词不仅受上下文影响,还会对其他文本和语境中字词的理解产生影响,它们产生的意义超越限定的文本和时空而纠缠在一起。同音字、谐音字、形近字、避讳字及常用的比喻、拟人、用典等修辞,会受后世的视角与解释影响,意义多歧并无限延异。理解诗歌文本整体性的视角、方法和情境不同,结论和意义也就各异。后世的注解无法改变历史文本,但会改变对历史文本及其意义的理解。古典诗歌有深厚的隐喻传统和象征体系,数字化诗歌文本中的互文性揭示,有望将这种被后世文学观念遮蔽的传统揭示出来。

互文性是相容的可观测量的最大集合来定义一个文本,对一个字词的解释就是一次暂且定义,会影响到这个字词在其他文本中的理解,如字面义、引申义、隐喻义等。互文性是古典诗歌突出的文本特征和普遍现象,王士禛《带经堂诗话》说“唐人佳句,多本六朝,昔人拈出甚多”②,甚至有“唐人诗皆自选出,改一二字便为己作,即李杜亦然”③的说法。每个字词的含义受上下文其他字词出现与否及排列方式的影响,数据库又将上下文极大拓展。非定域、映射纠缠和叠加态不仅让物体能完全无视空间的分离而相互连接,也产生了时间上反向因果的错觉,或者允许两个事件的因果顺序发生叠加。比喻、双关、拟

① 蒂费纳·萨莫瓦约,邵炜译:《互文性研究》,天津:天津人民出版社 2003 年版,第 82—83 页。

② 王士禛著,张宗柟纂集,夏闳校点,《带经堂诗话》卷十五,北京:人民文学出版社 1963 年版,第 400 页。

③ 骆鸿凯:《文选学》源流第三,北京:知识产权出版社 2012 年版,第 93 页。

象等构成诗歌之为诗歌的缘故,正是词义叠加和文本互文的体现。如果一首诗歌无法与其他文本发生关联想象,诗歌的意义就无法充分展开,甚至无法实现,也就失去了能动感发的诗意。

数据是证据的扩大和延伸,大数据带来更多证据的同时,也带来了细颗粒度和碎片化知识对齐难题,从而导致研究排除“对起源、原因、出处、影响与目的等问题的关注”,进而“对历史过程的任何目的论或因果关系的观点持激烈的敌视态度”①。数字化文本的非线性、整体性、关系性和过程性等特征,造成了“无角度,不学术”现象。在更大的数据整体中,学科也只是一种方法,每种研究都是整体的一个分支。不同专业的研究者把自己从事的特殊问题看成最重要的问题,忽视了对真正重要问题的关注,多是在定义、假设和同义语之间进行创造。结果只是提供某种情境下的可能性,而这种可能性能否实现,以及如何实现,取决于预设条件和技术路径。模型是对真实世界的抽象,抽象化和规范化才能可推导,产生具有可累积的知识。提升研究的维度和模型化思考,会让原本对立的观点能在更高层次和维度统一起来。

诗歌的字词解读受既有用例和用法的影响,从不同的用法和用例出发,会获得不同的理解与感受;不同角度的理解和认识在揭示不同隐义的同时,也不断地延展和扩散意义。诗体起源、创作背景、派别演变等文学史论述致力于发明节点的意义,但它们更是在过程中呈现的,后人的选择、比较与讨论赋予了以前诗作未具有的意义。诗歌的意义正是建立在读者与作者,经由文本和时代背景下的串联。所谓作者、文本与时代的统一,也是经由读者(研究者)的发现和赋予的。以 ChatGPT 为代表的语言大模型和生成式人工智能,让这种互文性和非定域性成为数字诗学的基本特征。

数字化让原本基于篇章、注引和整本书或学科领域的知识循环被打破,涌现了更多知识映射、学科破壁现象。一方面,数字化改变了文本、工具书和学科专业的阐释语境,拓展了诗歌的意义边界;另一方面,知识关联的层次、能力和场景,让诗歌文本从字词到篇章,都呈现出前所未有的开放状态。文献数字化让诗歌文本解读前所未有地依赖于语境,“言此意彼”“词不尽意”“心照非所宣”等让诗义具有即时性、丰富性和悖论特征。意义是所有可能理解的集合,基于不同的范畴和情景,它们之间甚至可能是矛盾的,但整体上又是互补和相生相成的。

余　论

数字化让世界呈现为文本方式,一切皆可运用数字表现和还原,并通过各种模型和算法来呈现和解释数据背后的现象和规律。当前基于字符匹配的全文检索,还未能释放数字文献应有的能量,如果把人名、地名、官职、典故等析取出来,赋予独立的标识符和实现通行的语义检索,而“一旦语言文字得到数字文本和结构化数据标准的支持,其‘可计算性’得到释放,或将在很大程度上使‘再造人文’成为现实”。② 我们应当走出辑佚式考证和集邮式研究,从系统层面来分析和认识对象整体。创建诗学分析模型和诗歌知识图谱,有助于在多面向和高维度审视研究对象,二维层面的交叉线,在三维立体层面可能并不相交。

印刷文本是二维结构,数字文献在通信协议和应用程序支持下,可以跨越不同平台,由单态语言转化为多模态语言。诗歌文献中的事实性知识如人物、时间、地点、典故等可以通过标引进行表示和聚合,概念性知识可以通过语义网络进行结构化和模式化。各种知识网络中,标引的层次和密度、节点的多少、边的方向和权重等都会让原有线性词句或二维句篇出现语义增强、知识增强的现象,以更好把握显性的符号知识和隐性的模型和向量知识。细颗粒、碎片化、深描特写和超文本聚合,让知识在原有体系

① 拉卡普拉、卡普兰主编,王加丰等译:《现代欧洲思想史——新评价和新视角》,北京:人民出版社 2014 年版,第 71 页。

② 刘炜、叶鹰:《数字人文的技术体系与理论结构探讨》,《中国图书馆学报》2017 年第 5 期,第 32—41 页。

中抽离出来,被充分关注和增强感知,甚至联觉表意和沉浸式体验。数字文本作为真实文本与阅读体验的中间层,从字词、句法到格律、体式,从意象、情感到用典、诗意,从聚类、拟像到交互、场景,算法、模型让我们习焉不察而起着越来越大的主导作用。

随着古籍数字化进程不断加快,诗歌文献形态和技术方法的变革,新的诗学范式呼之欲出,但累积开放的精良数据仍然有限,实体、属性的消歧和对齐尚处于初始阶段,文本分析工具还不先进,堪作示范的成功案例不多。即使引入人工智能也多为不透明的"黑箱"系统,往往是"利用深度学习算法和计算机的算力优势处理数量巨大且纷乱复杂的历史样本,从中得到一个预测模型,并以此处理与历史样本同质或相似的新样本"①。信息本身只能产生信息,信息需要接入人的回路才能导流为知识和经验,在人与人之间产生价值和意义。诗歌是在人参与下被激活,通过不断被阅读、注解和阐释而增值,正如回文诗可以被有经验的读者串读出多种意思,作诗软件生成的诗歌也可以被"读出"意思。文本是一种召唤结构,赋予其意义的是研读者的知识结构、情感体验与认识想象。就像测字、卜筮一样,是由人的先天结构与图形文字共同"合谋"而成。诗歌因为人的参与和连接而产生意义,正如古希腊哲学家普罗泰戈拉所说:"人是万物的尺度,是存在的事物存在的尺度,也是不存在的事物不存在的尺度。"②人文研究不能一直以其特殊性来规避普遍性,也不能因为普遍方法的引入,而忽略人的价值和人文立场。人虽然不是这个世界上的意义的全部,却是意义循环中的最关键一环。数字让诗学和人文更新,而人让数字和诗文有意义。

The Digitization of Ancient Texts and the Paradigm Innovation in Traditional Poetics

Li Feiyue

(*School of Humanities and Social Sciences*, *Tsinghua University*, *Beijing* 100084, *China*)

Abstract: The aggregation of various electronic poem texts has given rise to a vast collection of macrotexts and hypertexts. The advancement of technologies such as word segmentation, annotation, and semantic technique has further driven the development of poetic knowledge towards fine-grained and high-dimensional analysis. Databases, functioning as a form of text, have reset the classical order of poems and prompted a redefinition of poetic concepts, shifting from substances to relationships. Empirical research has made up for the shortage of standards and remedied the arbitrariness in interpretation, re-establishing poetics as a technical discipline and enhancing its scientific validity and legitimacy. Meanwhile, modelling and algorithms have brought to light latent variables, hidden knowledge and implicit contexts, highlighting the intertextuality and non-locality of texts, and underscoring the role of humans in the chain of meaning. The in-depth development of ancient textual digitization, knowledge enhancement, and multimodal large language models has led to a systematic transformation in poetics, from the reshaping of the classics, the reconstruction of concepts, and the restructuring of relationships to the re-evaluation of values.

Key words: digital humanity; poetic technique; intertextuality; paradigm

(责任编辑　郑　园)

① 李丰:《人工智能与艺术创作——人工智能能够取代艺术家吗?》,《现代哲学》2018 年第 6 期,第 95—100 页。

② 北京大学哲学系外国哲学史教研室编:《古希腊罗马哲学》,北京:三联书店 1957 年版,第 138 页。